

LOG-ANALYZING SYSTEM

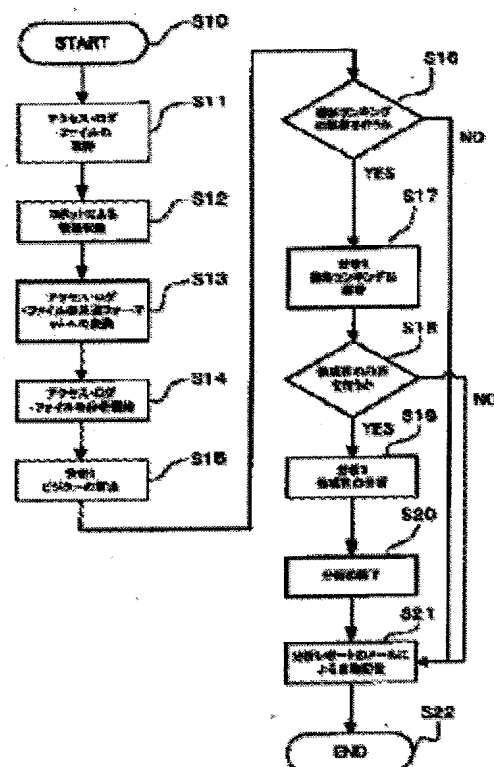
Publication number: JP2002063102
Publication date: 2002-02-28
Inventor: SAITO TAKASHI
Applicant: FRONTLINE DOT JP KK
Classification:
 - International: G06F13/00; G06F13/00; (IPC1-7): G06F13/00
 - European:
Application number: JP20000247126 20000816
Priority number(s): JP20000247126 20000816

Report a data error here

Abstract of JP2002063102

PROBLEM TO BE SOLVED: To easily calculate page transition ranking by calculating the estimated number of visitors from an access log. **SOLUTION:** In this system, estimated visitors are specified from the kind of an IP address and a Web browser and an OS, in which the browser is operated, and the number of visitors is analyzed by defining the persons who continuously perform access in a fixed time as the estimated visitors. Moreover, a site manager is able to estimate how many visitors are guided through pages, as scheduled by embedding a specific meta tag in a home page, or setting a home page which is a bench mark, and retrieving the page which is the bench mark by a robot retrieval engine in this log-analyzing system, and observing the bench mark transition by using the retrieved page as a pivot. Moreover, the means are combined, so that log data among plural servers can be analyzed efficiently.

ログ分析システム全体の流れ



Data supplied from the *esp@cenet* database - Worldwide

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号
特開2002-63102
(P2002-63102A)

(43) 公開日 平成14年2月28日 (2002.2.28)

(51) Int.Cl.⁷
G 0 6 F 13/00

識別記号
5 4 0

F I
G 0 6 F 13/00

テマコード* (参考)

5 4 0 R

審査請求 有 請求項の数13 O L (全 18 頁)

(21) 出願番号 特願2000-247126(P2000-247126)

(22) 出願日 平成12年8月16日 (2000.8.16)

(71) 出願人 300064847

株式会社フロントライン・ドット・ジェービー

東京都目黒区下目黒1丁目8番1号 アルコタワー5F

(72) 発明者 斉藤 隆史

東京都目黒区下目黒1丁目8番1号アルコタワー5F 株式会社フロントライン・ドット・ジェービー内

(74) 代理人 100106851

弁理士 野村 泰久

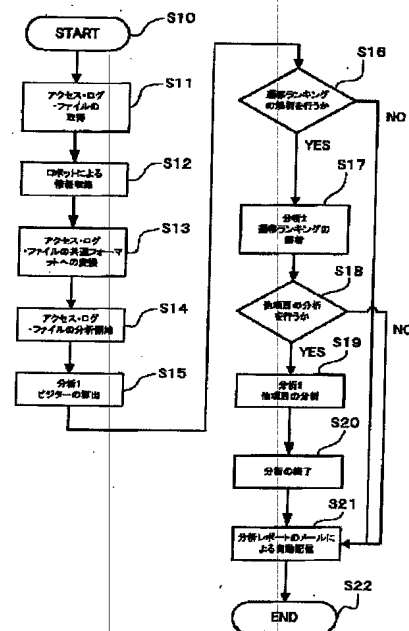
(54) 【発明の名称】 ログ解析システム

(57) 【要約】

【課題】 アクセス・ログから推定ビジター数を算出し、ページ遷移ランキングを簡易に算出すること。

【解決手段】 本発明は、IPアドレスとウェブブラウザの種類と、そのブラウザが動くOSにより推計ビジターを特定し、更に、一定時間のアクセスを継続的にしている人間を推定ビジターとして、ビジター数を分析するシステムである。さらに、サイトの運営者が特定のメタタグをホームページ上に埋め込みあるいはベンチマークとするホームページを設定することにより、ログ解析システムの中でロボット検索エンジンによりベンチマークとなるページを探り当て、それをピボットしてページ遷移を観察することにより、何人のビジターが所定通りのページを導かれたかを測定する。さらに、これらの手段を組み合わせることにより、複数のサーバー間のログデータを効率よく解析する。

ログ解析システム全体の流れ



【特許請求の範囲】

【請求項1】 所定のWEBサーバのアクセス・ログを取得し、該ログ・ファイルからユーザのWEBサーバへのアクセス動向を動的に解析するログ解析システムにおいて、

上記アクセス・ログのIPアドレス、ウェブブラウザとOSの種類の一貫性、及び各HTMLファイルのセッション間遷移の時間間隔に基づいてビジター数を推計することを特徴とするログ解析システム。

【請求項2】 前記セッション間遷移の時間間隔が30分以下の場合、同一ビジターであると判定することを特徴とする前記請求項1記載のログ解析システム。

【請求項3】 前記ウェブブラウザとOSの種類はバージョン情報を含むことを特徴とする前記請求項1または2記載のログ解析システム。

【請求項4】 特定のHTMLファイルを指定し、該特定のHTMLファイルをロボット検索手段によって前記ログ・ファイルから探し出し、該指定されたHTMLファイルの前もしくは後のHTMLファイル間の時系列遷移を解析することにより、ユーザのWEBサーバへのアクセス動向を動的に解析することを特徴とする前記請求項1～3の内、いずれか1項記載のログ解析システム。

【請求項5】 前記特定のHTMLファイルのヘッダー部分に事前に特定のメタ・タグを埋め込み、該メタ・タグをロボット検索手段によって前記ログ・ファイルから探し出すことにより、前記特定のHTMLファイルを指定することを特徴とする前記請求項4記載のログ解析システム。

【請求項6】 前記HTMLファイル間の時系列遷移の特定のパターンを有する前記ログを集計することを特徴とする前記請求項4または5記載のログ解析システム。

【請求項7】 前記HTMLファイル間の時系列遷移の特定のパターンと部分的に一致する前記ログを集計することを特徴とする前記請求項4または5記載のログ解析システム。

【請求項8】 前記所定のWEBサーバが複数存在し、該複数のWEBサーバの前記アクセス・ログを共通のフォーマットに変換して、ログ分析サイトのログDBに転送することを特徴とする前記請求項1記載のログ解析システム。

【請求項9】 前記複数のWEBサーバが異なる種類のサーバであって、前記ログDBにサーバ名と共にアクセス・ログを転送することを特徴とする前記請求項8記載のログ解析システム。

【請求項10】 前記複数のWEBサーバが同一種類のサーバであって、前記ログDBにアクセス・ログのみを転送することを特徴とする前記請求項8記載のログ解析システム。

【請求項11】 前記推計されたビジター数に基づき、アクセス数／ビジター数／参照時間を表示してレポートを

作成することを特徴とする前記請求項1～9の内、いずれか1項記載のログ解析システム。

【請求項12】 前記アクセス数／ビジター数／参照時間により、任意のパラメータ毎のランキングを表示してレポートを作成することを特徴とする前記請求項11記載のログ解析システム。

【請求項13】 前記解析された各HTMLファイルのヘッダー部分のタイトルをロボット検索手段によって取得し、該取得したタイトルを併せて表示してレポートを作成することを特徴とする前記請求項11または12記載のログ解析システム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】この発明は、インターネット上でのアクセスログを管理、分析のための運用システムに関する。

【0002】

【従来の技術】従来、インターネット上で訪問者を解析するシステムには、いろいろなシステムが提案されている。

【0003】訪問者数を測定ないしは推定するのには、インターネット通信で求められるブラウザに具備されているクッキーIDを用いたり、IPアドレスを用いるものが存在している。また、それらの訪問者がインターネット上のホームページの訪問をどのように行っているかに関しては、すべてのアクセスログを取得し、統計的な処理を行い、これらの訪問者のIDとを照らし合わせ、解析を行うということが行われている。さらに、アクセスログを効率よく処理しようとするため、「コレクタ」と呼ばれるものをルータの前面に設置し、わざわざログだけをパケットレベルで取得するハードウェアをかませることによって、効率よく訪問者数を把握し、くわえて、ホームページ上のページ訪問（ページ遷移）を計算する仕組みが採用されている。また、ロボット検索エンジンは、インターネットホームページの検索に際してよく使用されている。

【0004】しかしながら、訪問者数を推定する際に、例えば、クッキーIDについてみれば、プライバシー保護の観点から、自らクッキーID発行機能の停止を行うことにより、正確な人数を測ろうとするあまり、クッキーIDが捕捉できる人間のみの観測になってしまう欠点がある。また、サーバの中にはクッキーを埋め込む機能を有しないものも数多く存在し、この場合はクッキーによるアクセス・ログが取れない。

【0005】また、IPアドレスによる解析には、特に企業からのアクセスに対しては、トラフィックとトランザクションがファイヤーウォールを通過するためIPがファイヤーウォールのIPとなってしまう、大勢の人間がアクセスしているにもかかわらず、単一の人物と把握されてしまう可能性が大きかった。

【0006】また、特にページの遷移を分析する場合は、すべてのログを分析するのは膨大なデータを扱う上、集計の手間と分析の手間が多く、コンピューターリソースの面からも金銭的な面からも多大な費用を要していた。

【0007】さらに、WWWサーバーが複数になると、アクセスログデータの統合はさらに複雑になり、ログを統合する際に非常に多くの手間がかかる。これらの課題を効率よく解決する手段が求められていた。

【0008】

【発明が解決しようとする課題】この発明は、アクセスしているビジターを簡単に正確に同一人物と推定する手法を提供し、また該ビジターのページ間遷移を容易に把握できる手法を提供することを目的とする。

【0009】

【課題を解決するための手段】この発明は、以上のような課題に対して、IPアドレスとウェブブラウザの種類と、そのブラウザが動くOSにより推計ビジターを特定し、更に、一定時間のアクセスを継続的にを行っている人間を同一人物と推定として、ビジター数を分析するシステムである。

【0010】さらに、サイトの運営者が特定のメタタグをホームページ上に埋め込みあるいはベンチマークとするホームページを設定することにより、ログ解析システムの中でロボット検索エンジンにより、ベンチマークとなるページを探り当て、それをピボットして、前後のページ遷移を観察することにより、何人の訪問者が、所定通りのページを導かれたかを測定する。また、これらの手段を組み合わせることにより、複数のサーバー間のログデータを効率よく解析する。

【0011】すなわち、この発明のログ解析システムは、所定のWEBサーバのアクセス・ログを取得し、該ログ・ファイルからユーザのWEBサーバへのアクセス動向を動的に解析するログ解析システムにおいて、上記アクセス・ログのIPアドレス、ウェブブラウザとOSの種類の同一性、及び各HTMLファイルのセッション間遷移の時間間隔に基づいてビジター数を推計することを特徴とする。

【0012】この場合、前記セッション間遷移の時間間隔が30分以下の場合、同一ビジターであると判定し、前記ウェブブラウザとOSの種類はバージョン情報を含むことを特徴とする。

【0013】また、この発明のログ解析システムは、特定のHTMLファイルを指定し、該特定のHTMLファイルをロボット検索手段によって前記ログ・ファイルから探し出し、該指定されたHTMLファイルの前もしくは後のHTMLファイル間の時系列遷移を解析することにより、ユーザのWEBサーバへのアクセス動向を動的に解析することを特徴とする。

【0014】さらに、前記特定のHTMLファイルのへ

ッダー部分に事前に特定のメタ・タグを埋め込み、該メタ・タグをロボット検索手段によって前記ログ・ファイルから探し出すことにより、前記特定のHTMLファイルを指定することを特徴とする。

【0015】さらに、前記HTMLファイル間の時系列遷移の特定のパターンを有する前記ログ、あるいは該時系列遷移の特定のパターンと部分的に一致する前記ログを集計することを特徴とする。

【0016】また、この発明のログ解析システムは、前記所定のWEBサーバが複数存在し、該複数のWEBサーバの前記アクセス・ログを共通のフォーマットに変換して、ログ分析サイトのログDBに転送することを特徴とする。

【0017】この場合、前記複数のWEBサーバが異なる種類のサーバであって、前記ログDBにサーバ名と共にアクセス・ログを転送する場合と、前記複数のWEBサーバが同一種類のサーバであって、前記ログDBにアクセス・ログのみを転送する場合が存在する。

【0018】最終的に、この発明のログ解析システムは、前記推計されたビジター数に基づき、アクセス数／ビジター数／参照時間を表示してレポートを作成することを特徴とし、前記アクセス数／ビジター数／参照時間により、任意のパラメータ毎のランキングを表示してレポートを作成したり、前記解析された各HTMLファイルのヘッダー部分のタイトルをロボット検索手段によって取得し、該取得したタイトルを併せて表示してレポートを作成することを特徴とする。

【0019】

【発明の実施の形態】以下、図面を参照しながら、この発明の好適な実施の形態について詳細に説明する。

【0020】本発明の実施の形態の一例は、以下のとおりである。図1に、この発明に用いられるネットワークの全体図を示す。図1において、1はインターネットであって、クライアントサイトのWWWサーバ10がルータ2を介してログ分析サイトに接続される。ログ分析サイト側には、レポート配信用MAILサーバ3、WWWサーバ4、情報処理サーバ5、ログ分析サーバ6、ロボット検索サーバ7、ログDB8、ロケーションDB9、クライアントデータベース15が接続される。上記のように、ログ分析サイトは、個々のクライアントサイトの生ログDB11にアクセスすることにより機能する。なお、このサーバー群においては、周知のOSが具備され、この上でログ解析システムのプログラムが動作する。

【0021】図2は、この発明のログ解析システムにおけるログ分析サイト全体の流れを示すフローチャートである。このログ解析サイトが動作を開始する（例えば、一日一回）と、ログ収集サーバ12によってクライアントサイトのアクセスログのが収集される（S11）。次に、S12において、ロボット検索サーバ7によって、ク

クライアント上におけるどのホームページを表示するファイルが、分析の対象となるかデータの収集を始める。S13においては、ログ収集サーバ12によって、アクセス・ログをログ分析サーバ6によって解析するための共通フォーマット(図10)への変換が行われる。この共通フォーマット化へは、ログファイルにおいて主に3種類のフォーマット(Netscape(登録商標)、NCSA、IIS)が混在しているためである。例として、図9には時刻の統一をする際のプロセスが開示されている。S14において、アクセス・ログの分析が開始される。S15においては、始めにビジターの算出が行われ、S16においてビジターの推計のみで終了する場合には、S21へと移行する。S16において、ホームページの遷移のランキングの解析を行う場合には、S17において、遷移ランキングの解析が行われ、S18へと移行する。S18においては、他の項目の分析を行うかどうかの指示を行い、他項目の分析を行う場合には、S19に移行し、そうでない場合には、S21に移行する。S19において他項目の分析を行った後、S21において、分析レポートを自動的に各クライアントに送付を行い、全体の解析を終了させる。これらはすべて所定のフォーマットに格納されるようになっており、図14のようなフォーマットが用意されている。かくして、周知のDBを用いてデータが格納される。

【0022】図3は、ログ・ファイルの取得から分析処理実行前までの詳細なステップを表している。S30においてプログラムが作動すると、S31でクライアント・サイトへアクセスし、S32に移行する。S32において、クライアント・サイトで保存されている生ログDB11にアクセスを行う。アクセスを行うと同時に、S33において、生ログファイルの取得方法はFTP(インターネットで使用するFile Transfer Protocol)によってどのように取得するかが判断され、FTPにおいてGET、すなわちログ分析サイト側から生ログを取得する場合には、S34に移行し、ログ・ファイルがあるかどうかを確認した後、S35において、ログの日次ログ・ファイルがあるかどうか判断され、日次ファイルがない場合には自動的に一定時間にS34へ戻り、処理を再実行する。日次ログファイルが存在する場合、S36に移行し、ログ・ファイルを取得する。ここで、S33のときファイルの取得方法がFTPのPUT、すなわちクライアントサイト側から定時にファイルが転送される場合には、そのままログ・ファイルの取得S36に移行する。その後、S37においてはログ・ファイルの共通フォーマットに変換され(図10)、このファイル一つ一つにIDを付加して、DB(S38)に登録し、登録終了後終了処理を行う(S39)。

【0023】図4は、図3のS38で登録された、データが、どのように解析されるかを示している。S41において処理がスタートされると、S42において、データの並

び替えを行う。ここでは、ID(レコード)ごとに記録されているDBをソース・アドレス、使用ブラウザ(バージョン情報を含む。)、アクセス日時、URLの項目順に並び替えを行いビジター算出ステップに移行する。S44においてはまず、前後のレコードのソース・アドレスが同一かどうか判定され、(図10の表を参照)、同一であれば、前後のレコードの使用ブラウザが同一かが判断される(S45)。さらに同一であれば、前後のレコードのアクセス日時が前後30分以内かが判断されて(S46)、S44、S45、S46が同一であれば、同一ビジターと判断され、S49に移行する。S44、S45、S46のいずれかの段階で同一でない場合は、別ビジターとして登録され(S48)、S48とS49を合わせて、ビジターの算出が終了し(S49)、ビジター算出のブロックが終了する(S50)。なお、上記30分という同一ビジターであることの判定時間は、ページの中味によっても変動し、15〜45分の適当な時間に設定すればよい。なお、該判定時間は、別途統計を取って最適値に設定することが好ましい。

【0024】これをレコード分繰り返すことにより、当該ログの訪問者の人数の算出が図11のように可能となる。このときログDB8にはビジターごとにIDが振られ、ログが管理される。

【0025】次に図5において、ロボット検索サーバ7における、遷移の対象となるホーム・ページのファイルの検索方法を説明する。情報収集が開始されると(S58)、各サイトのトップディレクトリにロボットがアクセスを開始する(S59)。次に、S60において、ファイルソース(HTMLソースで書かれているもの)から情報収集を行い、S61においてタイトルが表記されているかどうか判断を行う。ここでタイトルが獲得できれば、S62に移行し、ロケーションDB9に格納する。タイトルタグが存在しない場合は、S63に移行する。続いて、S63において、アクセスしたファイルのソースにメタタグが記述されているか判断を行う。これは、事前にクライアント側で登録される、指定ページ用のメタタグの記述方法は以下のようにになっている。

【0026】・遷移を記録する最初のページには、
 <META NAME="flog-anchor" CONTENT="trans-top">

・遷移を記録する最後のページには
 <META NAME="flog-anchor" CONTENT="trans-end">
 のように記述されている。

【0027】ここで、上記あるいは図13のようなメタタグがソースに記述されていると、遷移ランキングの指定ページとして情報をロケーションDB9に格納する(S64)。その後、S65において、ソースにリンクの情報が存在するかが判断され、存在すれば、S66に移行し、リンク先のページに移動し、S72からのステップが実行される。S65において、ソースのリンク情報が記述されていない場合には、そのままS67に移行し、ロボッ

ト検索を終了させる。

【0028】図4で導き出された推定ビジター数と図5で得られた遷移の対象となるホーム・ページの位置に基づいて、図6において、ホーム・ページの遷移のランキングが算出される。ここで、ランキングの算出がスタートされると(S70)、推定ビジターのデータが取りこまれて(S71)、ビジターIDごとの分析を開始する(S72)。S73において、ビジターIDごとに図5において収集を行った指定ページが含まれているデータがあるかどうか判断され、データがある場合には、ログDB8に遷移ランキングの対象データとして、データが保存される。S73において対象データが存在しない場合には、S75において、次のビジターIDのデータがあるかが判断され、存在する場合にはS72に戻る。S75において、分析するレコードが存在しない場合には、S76においてDB格納されたデータにパターンごとに重複数が算出され、S77において指定順位のランキングが図12のように表示され、遷移ランキングの算出が終了する(S78)。

【0029】この遷移ランキングには2つの種類があり、

1) 指定したページからの遷移ランキング
2) 指定したページまでの遷移ランキング

を算出することができ、1)は指定したページからどのページに遷移していったかが把握され、また2)は指定したページまでのどのページから遷移してきたかが把握される。また、図12の実施例では遷移ページ数が3階層で、指定ページから3ページ先の出力ランキング数が表示される。

【0030】さらに、図7においては、他の項目における分析を行うフローチャートである。ここでは、図6で使用方法を利用して、以下のようなランキングを算出することが可能となっている。

【0031】

- 1 総アクセス数/総ビジター数/総参照時間
- 2 曜日別(日別)アクセス数/ビジター数/参照時間
- 3 時間帯別アクセスランキング
- 4 コンテンツ別アクセスランキング
- 5 ディレクトリ別アクセスランキング
- 6 サブドメイン別アクセスランキング
- 7 フルドメイン別アクセスランキング
- 8 ブラウザ別アクセスランキング
- 9 OS別アクセスランキング
- 10 最初のページ別ランキング
- 11 最後のページ別ランキング
- 12 前ページ(ファイル別)ランキング
- 13 前ページ(フルドメイン別)ランキング
- 14 検索キーワード別ランキング
- 15 検索エンジン別ランキング
- 16 ファイル転送量

- 17 エラーログランキング
- 18 時間帯別ビジターランキング
- 19 コンテンツ別ビジターランキング
- 20 ディレクトリ別ビジターランキング
- 21 サブドメイン別ビジターランキング
- 22 フルドメイン別ビジターランキング
- 23 ブラウザ別ビジターランキング
- 24 OS別ビジターランキング
- 25 最初のページ別ビジターランキング
- 26 総参照時間
- 27 コンテンツ別参照時間ランキング
- 28 ディレクトリ別参照時間ランキング
- 29 サブドメイン別参照時間ランキング
- 30 フルドメイン別参照時間ランキング
- 31 ブラウザ別参照時間ランキング

加えて図8においては、クライアント・サイトが複数にまたがっているときのログの解析処理を表現されている。ここでは、S90において処理が開始されると、S91においてログ・ファイルの取得が開始される。S92において、マージないしはミラー処理対象レポートかが判断される。ここでいう「マージ」は同一レポートの集計対象ログが複数にわたる場合を指し、単純にログを統合すれば良いもの、例えばオンライン・ショッピングにおける情報提供サーバーと決済サーバーなどが隔離されている場合に利用する。ミラーは、同一レポートの集計対象ログが複数にわたり、それを同一視して集計する場合を指す。この場合にはオンライン・ショッピングにおいて、同一アドレスにおいて処理の負荷分散を行うため複数台同じコンテンツが格納されているサーバーが接続されている場合を意味する。続いて、S92において処理を行う場合、S93においてサーバ情報を確認し、S94において、取得したログ・ファイルをそれぞれ分析処理を行う。S95において、マージ処理かミラー処理かの判断がなされ、マージ処理の場合にはS96に移行し、ミラー処理の場合にはS98に移行する。S96においては、それぞれの分析処理を同一レポートで別々のものとして出力し、S97へと移行する。S98においては、それぞれの分析処理を同一レポートで同一視して出力し、S97に移行する。また、S92において、処理対象レポートに当たらない場合には、S99に移行し、取得したファイルを分析処理し、S100の通常出力に移行する。S97において出力処理が終了の後、このプロセスが終了する。

【0032】また、この実施例では、ロボットによりメタタグを探し出す方法が開示されているが、これに限られることなく、メタタグを使用することなく特定のページをクライアント・サイトからでもログ解析システム側でも指定し同様の方法で解析することが可能である。

【0033】

【発明の効果】この発明によれば、今までIPアドレスやブラウザーのクッキーIDにより算出していたホーム

・ページへのビジター数の算定を、データが取得できないゆえに算出できなかった欠点を無くし、ビジター人数を、IPアドレスとウェブブラウザの種類と、そのブラウザが動くOSにより推計し、より正確にビジターを推定することが可能になる。

【0034】また、ページ遷移間の時間間隔によって同一ビジターと推定しているので、同一ビジターのアクセスを複数ビジターのアクセスと計数する間違いが減り、また他人のと他のアクセスを同一ビジターのアクセスと計数する間違いも減少する。

【0035】さらに、ロボット検索システムによって容易に指定ファイルを検索し、その情報と推定ビジター数を用いることによって、今までのように全ログファイルを解析することなく、ビジター数やページの遷移すなわちホーム・ページ上のビジターの動線を把握することが可能になるので、リソースの節約になりシステムの利便性が高い。加えて、今までのパッケージソフト類では不可能であった複数のログの統合された統計処理をスムーズに行うことができ、費用と工数の軽減になる。

【図面の簡単な説明】

【図1】この発明に用いられるネットワークの全体図を示す図である。

【図2】この発明のログ解析システム全体動作の流れを示す図である。

【図3】ログファイルの取得から分析処理実行前までの詳細なステップを表しているフロー・チャートである。

【図4】ビジター（推定訪問者）を算出するフロー・チャートである。

【図5】ロボット検索エンジンによる情報収集の流れを示すフロー・チャートである。

【図6】図4で得られた推定ビジター数を用いた遷移ランキングを算出するフロー・チャートである。

*

* 【図7】他項目における分析のフロー・チャートである。

【図8】複数サーバーのログを用いた解析例を示すフロー・チャートである。

【図9】中間ファイルの遷移を示す図である。

【図10】共通形式化された中間ファイルの一例を示す図である。

【図11】ログ分析レポートの出力サンプルを示す図である。

10 【図12】遷移ランキングを示す表示および出力サンプルを示す図である。

【図13】ロボット検索に使用するメタタグを埋め込んだ一例を示す図である。

【図14】このログ解析システムで用いるデータベースの概要を示す図である。

【符号の説明】

- 1 インターネット
- 2 ルーター
- 3 レポート配信用MAILサーバ
- 20 4 WWWサーバ
- 5 情報処理サーバ
- 6 ログ分析サーバ
- 7 ロボット検索サーバ7
- 8 ログDB
- 9 ロケーションDB
- 10 WWWサーバ（クライアント）
- 11 生ログDB
- 12 ログ収集サーバ
- 13 ファイアウォール
- 30 14 LAN
- 15 クライアントDB

【図13】

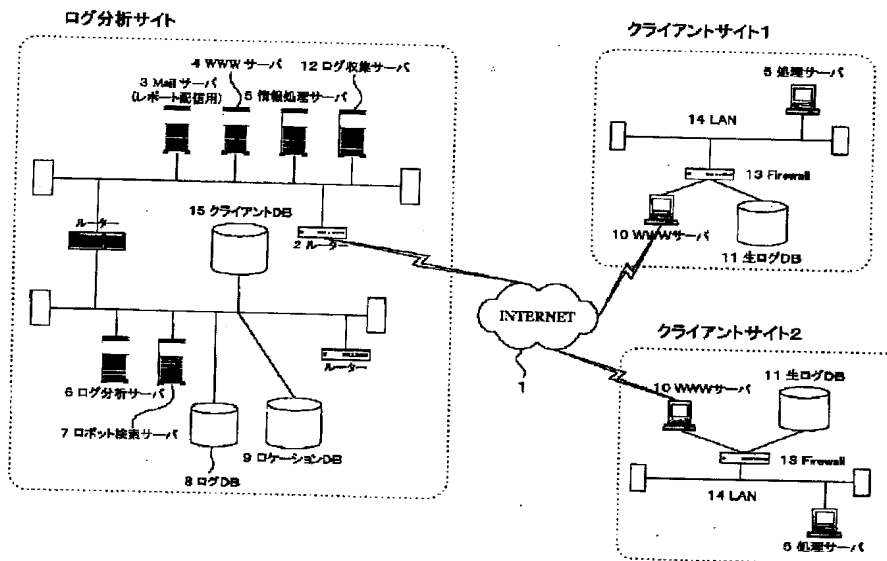
```

— 例 —
<html>
<head>
<title>Log Analysis Service</title>
<META NAME="log-anchor" CONTENT="trans-top">
<meta http-equiv="Content-Type" content="text/html; charset=x-sjis">
</head>

```

<<==== [ログ分析システムV1.1用TAG]

【図1】



【図9】

【中間ファイル 1】

入力ファイル中の該当行について、日付・時間情報を JST に換算し、"YYYY/MM/DD HH:MM:SS"形式に、TAB を 1個のスペースに置き換えて出力したもの。その他の項目については、元の形式のままで出力する。

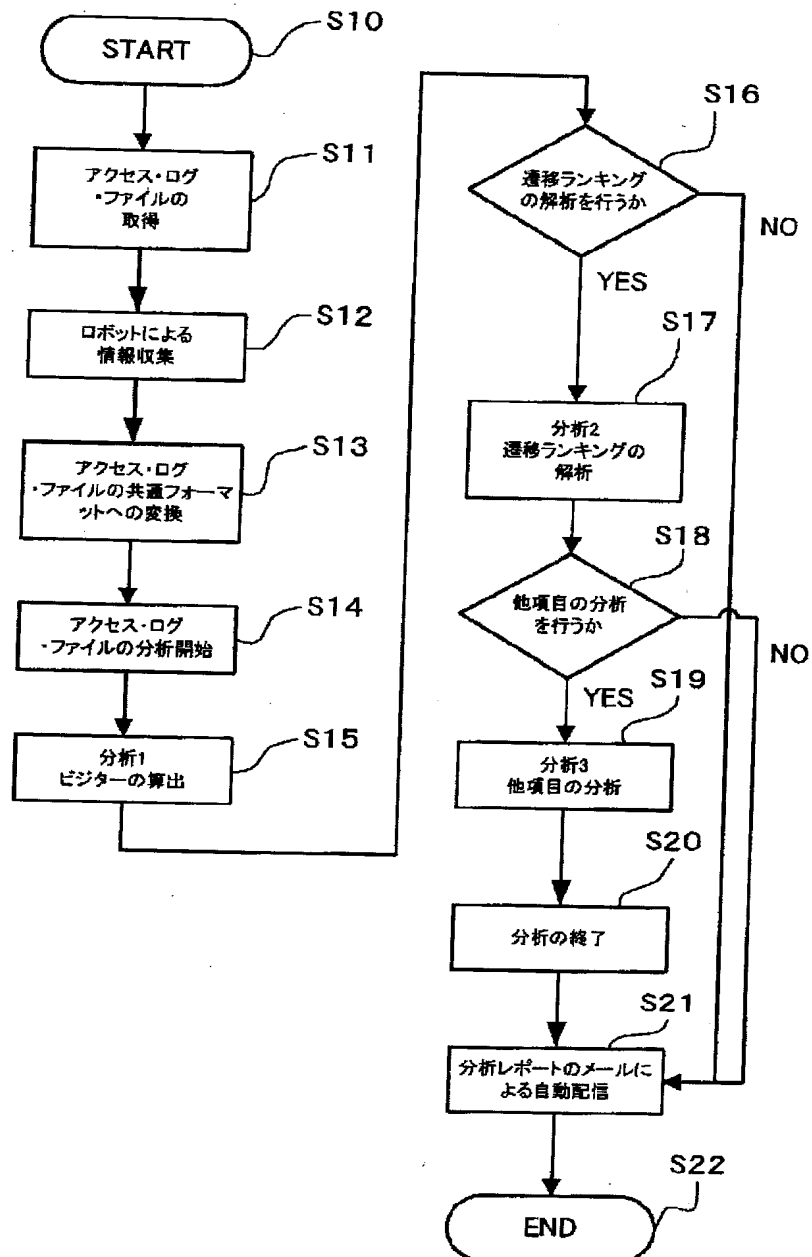
また、「#」「format=」で始まる行は、そのまま出力する。

<変換例>

```
(apache)
(略) kwgcc-01p52.ppp.odn.ad.jp -- [29/Jul/1999:00:00:00 +0900] "GET /POWERAXEL/image/s....
      ↓
      kwgcc-01p52.ppp.odn.ad.jp -- [1999/07/29 00:00:00] "GET /POWERAXEL/image/s.... (略)
(iiis3)
(略) 202.219.168.66, -, 99/06/26, 0:00:00, W3SVC1, YBPASP2, 202.221.4.99, 5406, 247, 13....
      ↓
      202.219.168.66, -, [1999/06/26 00:00:00], W3SVC1, YBPASP2, 202.221.4.99, 5406, 247, ..
.. (略)
(iiis4) (iis4 では GMT が使用されているものと仮定する。また、日付が省略されているので、
      15:00:00 ~ 23:59:59 を前日分、00:00:00 ~ 14:59:59 を当日分と仮定する)
(略) 00:00:02 210.131.44.143 - GET /olv/form_q.asp - 200 19641 Mozilla/3.0+WebTV/2.2.1+....
      ↓
      [1999/07/26 09:00:02] 210.131.44.143 - GET /olv/form_q.asp - 200 19641 Mozilla/3.0+W..
.. (略)
(netscape, time_diff=60 の場合)
(略) 210.131.71.98 -- [12/Jun/1999:07:19:10 +0900] "GET /uc_checkproc/imgs/topcalcostb....
      ↓
      210.131.71.98 -- [1999/07/12/ 08:19:10] "GET /uc_checkproc/imgs/topcalcostb.... (略)
(wu-ftp, time_diff=-60 の場合)
(略) Mon Aug 23 16:58:21 1999 935 202.211.205.166 3393630 /19990129/import/MAX/990823b.....
      ↓
      [1999/08/23 15:58:21] 935 202.211.205.166 3393630 /19990129/import/MAX/990823b..... (略)
```

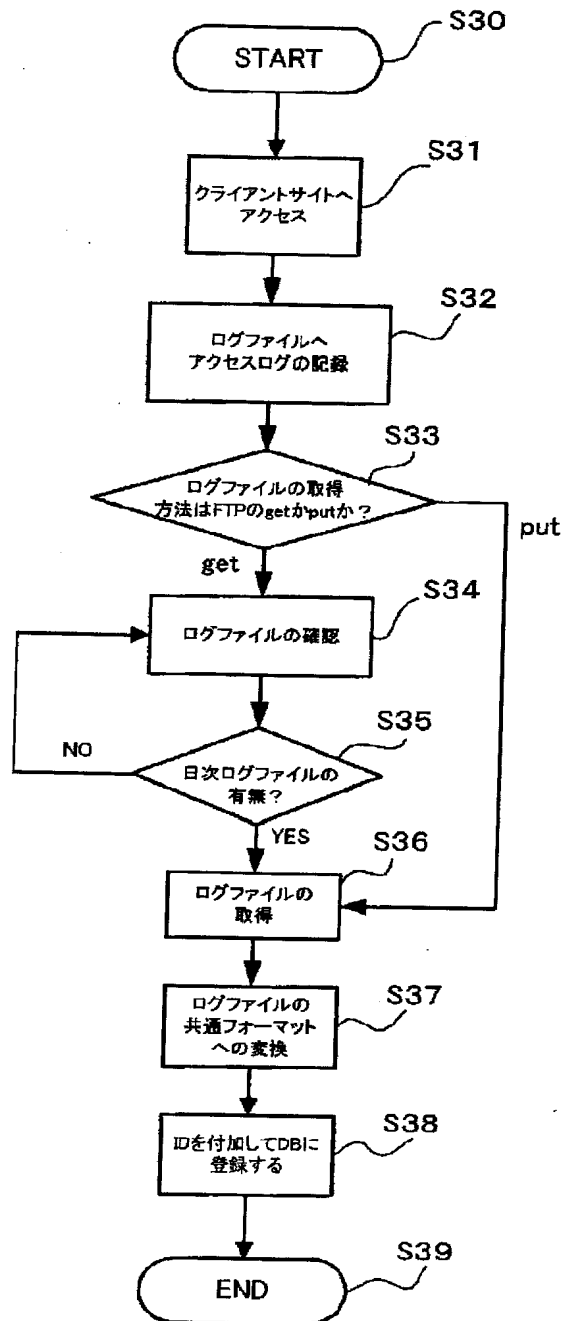

【図2】

ログ分析システム全体の流れ



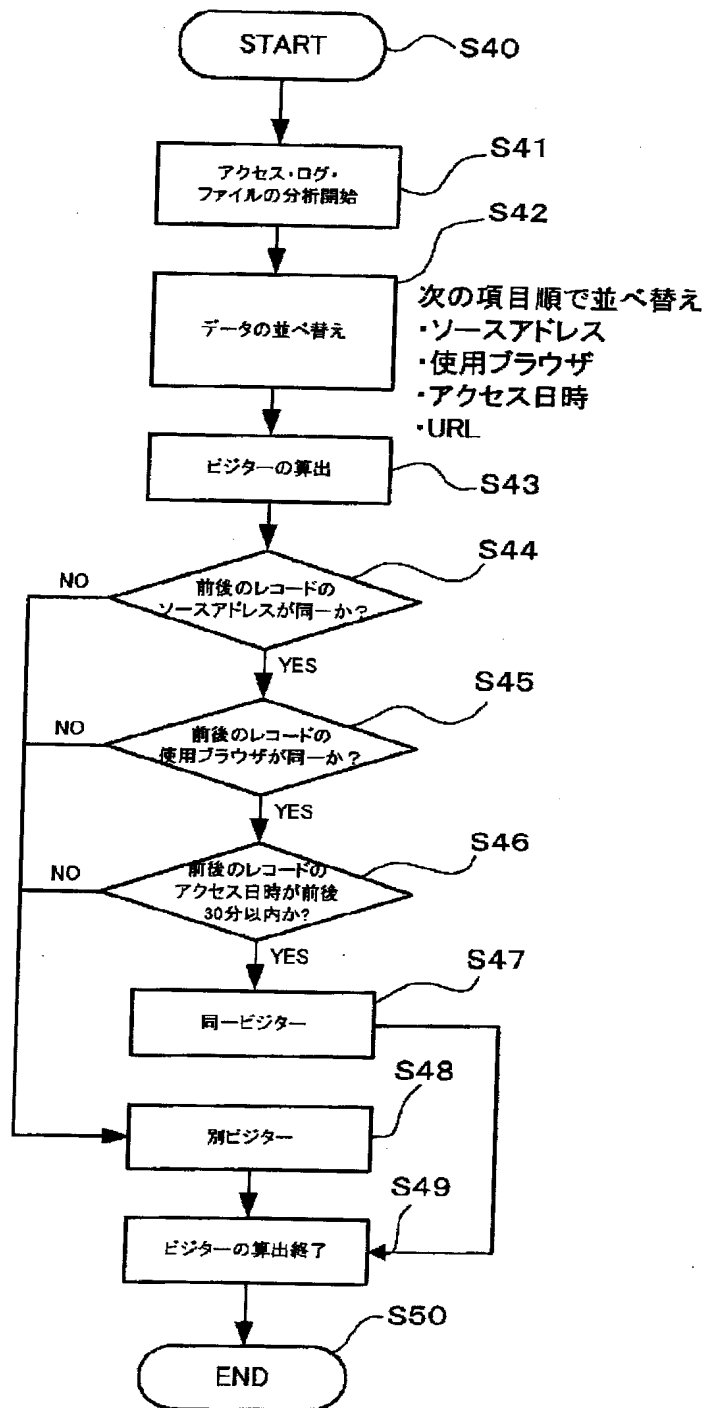
【図3】

ログファイル取得から 分析処理実行前までの流れ



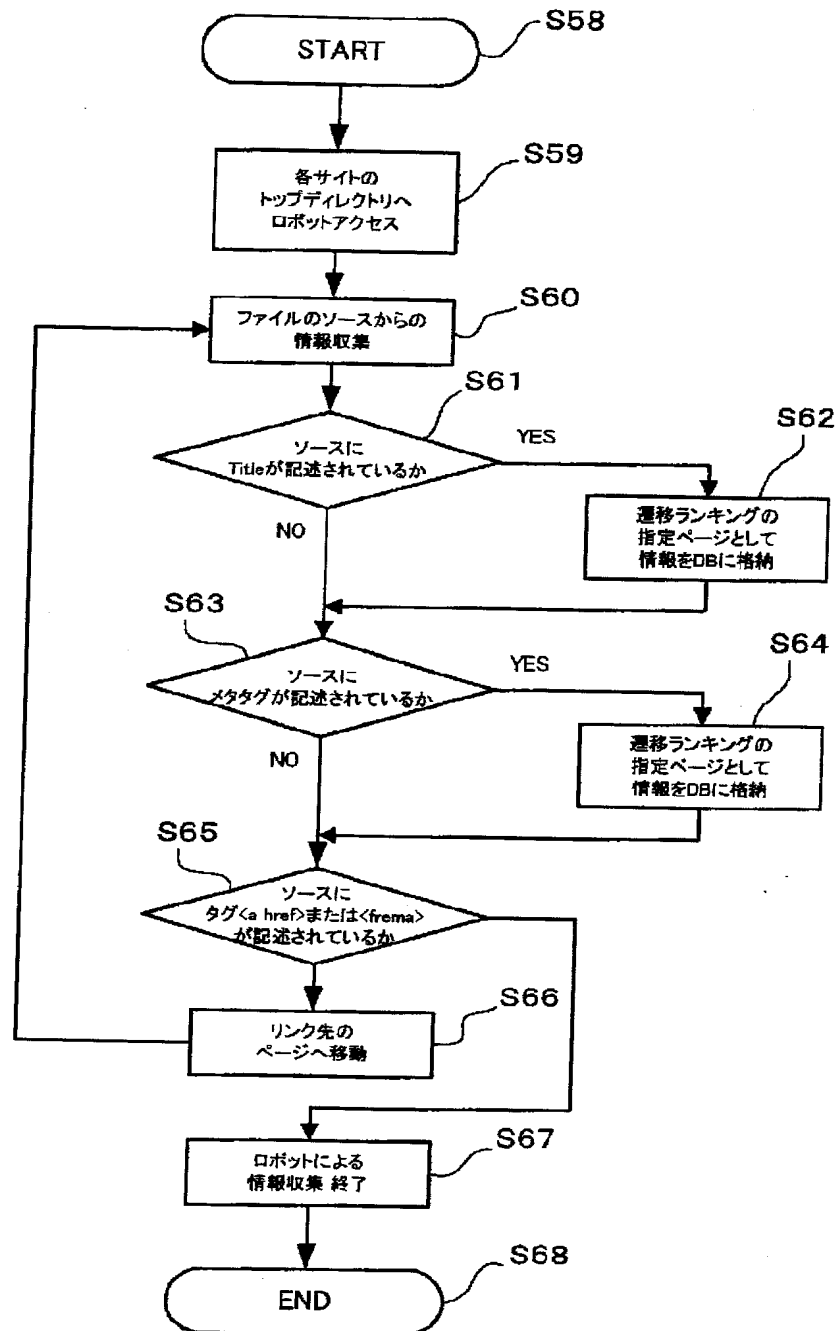
【図4】

ログ分析処理の流れ 1 ビジターの算出



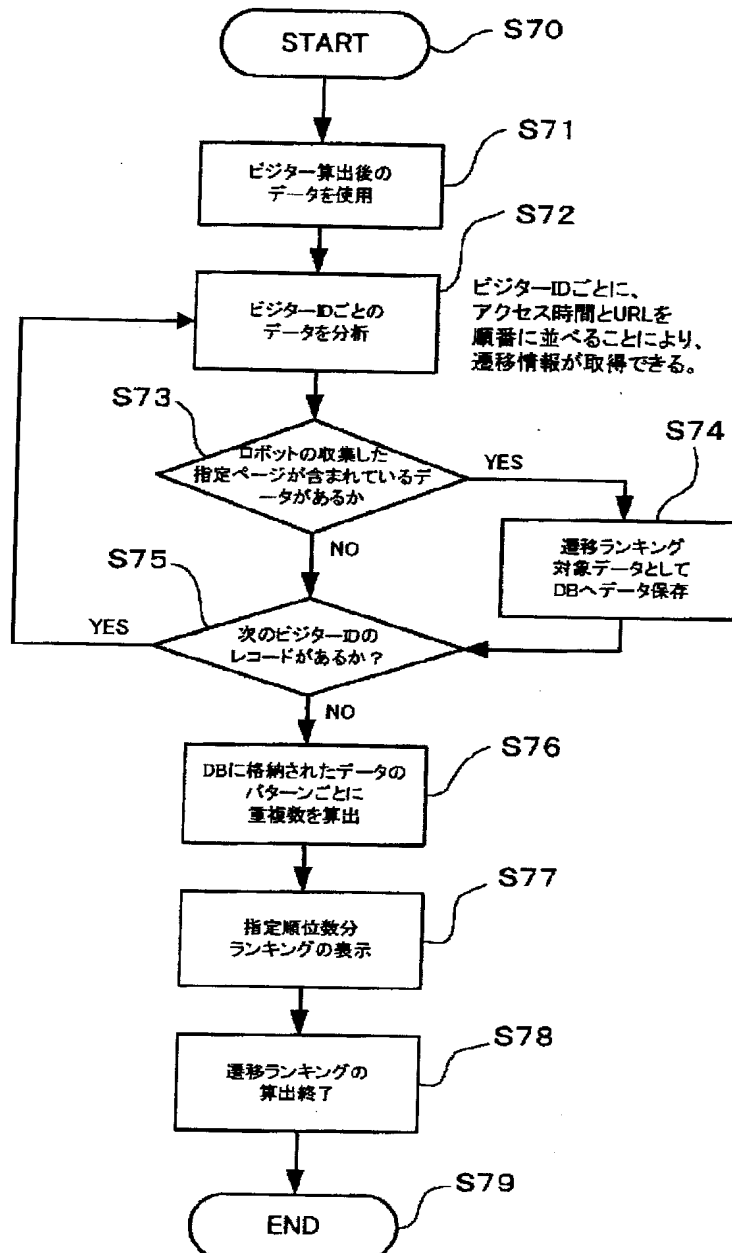
【図5】

ロボットによる情報収集の流れ



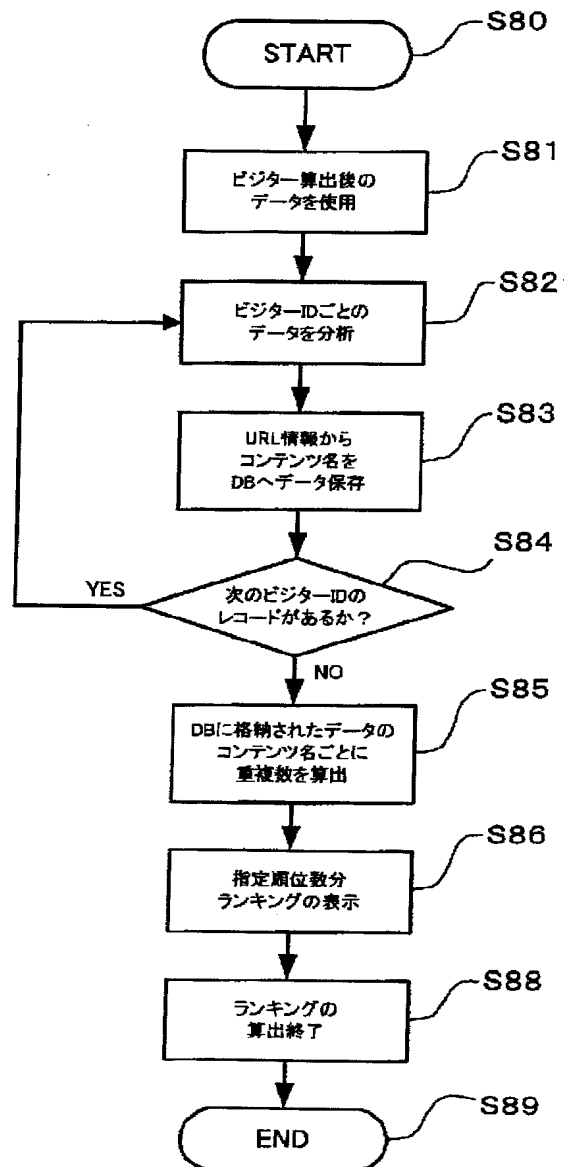
【図6】

ログ分析処理の流れ 1 遷移ランキング



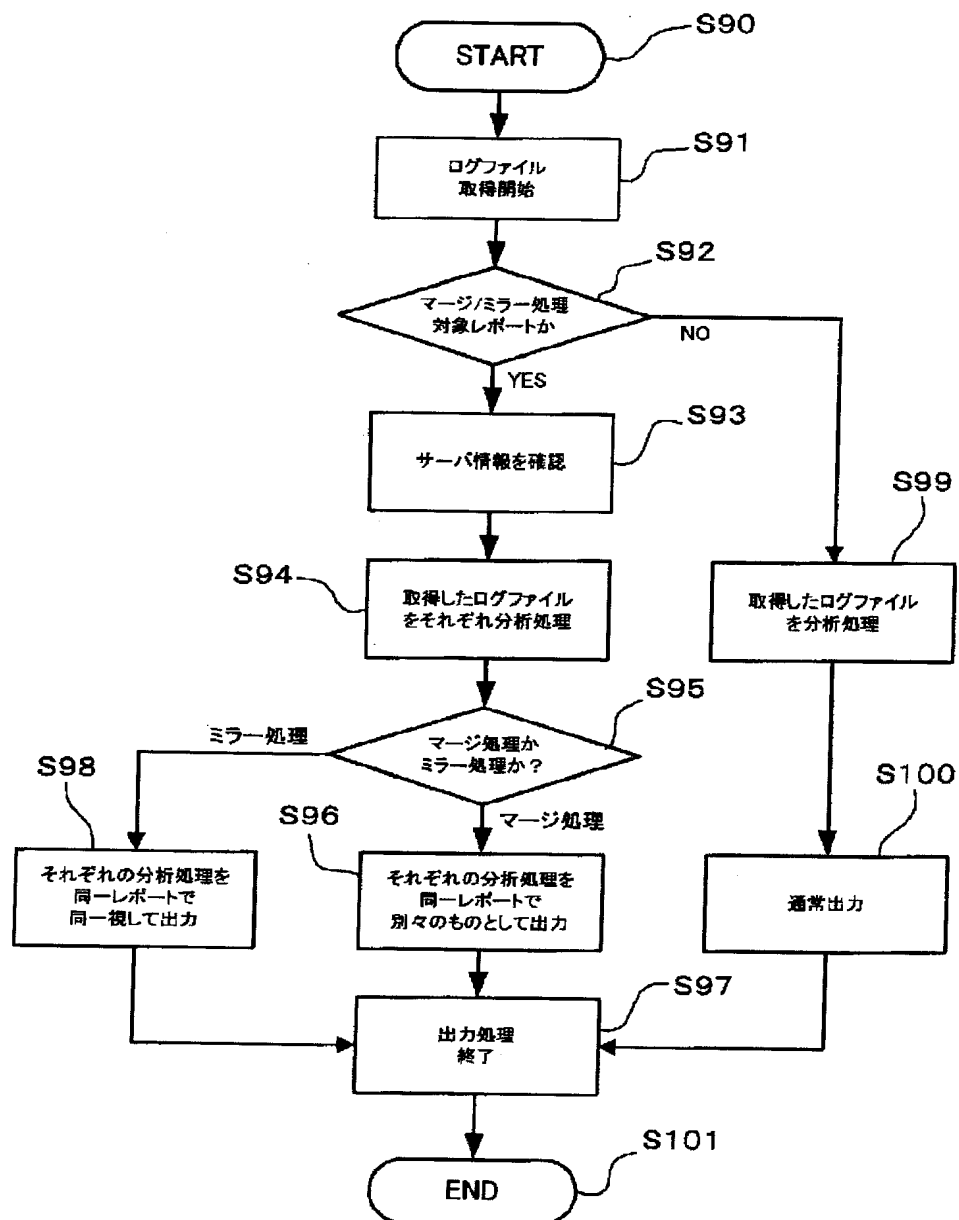
【図7】

ログ分析処理の流れ
他項目における分析例
[コンテンツ別ビジターランキング]



【図8】

ログ分析処理の流れ 複数ログファイルのマージ/ミラー処理



[中層ファイル]
 共通形式化され、
 一連のファイルが
 使用される。この
 とき、ファイル名
 をキーとして指定されたデータサンプル

※ 理解しやすくするため、1レコード毎に空白行を添えています
 実際には空白行はありません

2000/07/18 08:31:54	200 493	133.217.51.93	GET	/index1.htm	HTTP/1.1	Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)	http://www.fjj.co.jp/sevi.html
Windows 98; DigExt							
2000/07/18 08:31:54	200 2895	133.217.51.93	GET	/main2.html	HTTP/1.1	Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)	http://www.fjj.co.jp/index2.html
Windows 98; DigExt							
2000/07/18 08:31:55	200 87671	133.217.51.93	GET	/f12.gvt	HTTP/1.1	Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)	http://www.fjj.co.jp/index2.html
Windows 98; DigExt							
2000/07/18 08:31:58	200 405	133.217.51.93	GET	/images/copyright.gif	HTTP/1.1	Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)	http://www.fjj.co.jp/service/main.html
Windows 98; DigExt							
2000/07/18 08:31:58	200 556	133.217.51.93	GET	/images/main/title_servi	HTTP/1.1	Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)	http://www.fjj.co.jp/service/main.html
Windows 98; DigExt							
2000/07/18 08:31:58	200 160	133.217.51.93	GET	/service/images/bg.gif	HTTP/1.1	Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)	http://www.fjj.co.jp/service/main.html
Windows 98; DigExt							
2000/07/18 08:31:58	200 1370	133.217.51.93	GET	/service/images/head01	HTTP/1.1	Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)	http://www.fjj.co.jp/service/main.html
Windows 98; DigExt							
2000/07/18 08:31:58	200 2959	133.217.51.93	GET	/service/images/head02	HTTP/1.1	Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)	http://www.fjj.co.jp/service/main.html
Windows 98; DigExt							
2000/07/18 08:31:58	200 28770	133.217.51.93	GET	/service/images/service	HTTP/1.1	Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)	http://www.fjj.co.jp/service/main.html
Windows 98; DigExt							
2000/07/18 08:31:58	200 1016	133.217.51.93	GET	/service/images/service	HTTP/1.1	Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)	http://www.fjj.co.jp/service/main.html
Windows 98; DigExt							
2000/07/18 08:31:58	200 803	133.217.51.93	GET	/service/images/service	HTTP/1.1	Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)	http://www.fjj.co.jp/service/main.html
Windows 98; DigExt							
2000/07/18 08:31:58	200 714	133.217.51.93	GET	/service/images/service	HTTP/1.1	Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)	http://www.fjj.co.jp/service/main.html
Windows 98; DigExt							
2000/07/18 08:31:58	200 303	133.217.51.93	GET	/service/images/service	HTTP/1.1	Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)	http://www.fjj.co.jp/service/main.html
Windows 98; DigExt							
2000/07/18 08:31:58	200 404	133.217.51.93	GET	/service/images/service	HTTP/1.1	Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)	http://www.fjj.co.jp/service/main.html
Windows 98; DigExt							
2000/07/18 08:31:58	200 306	133.217.51.93	GET	/service/images/service	HTTP/1.1	Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)	http://www.fjj.co.jp/service/main.html
Windows 98; DigExt							
2000/07/18 08:31:58	200 303	133.217.51.93	GET	/service/images/service	HTTP/1.1	Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)	http://www.fjj.co.jp/service/main.html
Windows 98; DigExt							
2000/07/18 08:31:58	200 306	133.217.51.93	GET	/service/images/service	HTTP/1.1	Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)	http://www.fjj.co.jp/service/main.html
Windows 98; DigExt							
2000/07/18 08:31:58	200 303	133.217.51.93	GET	/service/images/service	HTTP/1.1	Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)	http://www.fjj.co.jp/service/main.html
Windows 98; DigExt							
2000/07/18 08:31:58	200 306	133.217.51.93	GET	/service/images/service	HTTP/1.1	Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)	http://www.fjj.co.jp/service/main.html
Windows 98; DigExt							
2000/07/18 08:31:58	200 303	133.217.51.93	GET	/service/images/service	HTTP/1.1	Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)	http://www.fjj.co.jp/service/main.html
Windows 98; DigExt							
2000/07/18 08:31:58	200 306	133.217.51.93	GET	/service/images/service	HTTP/1.1	Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)	http://www.fjj.co.jp/service/main.html
Windows 98; DigExt							

【図11】

< ビジター数 >

< ホームページ全体に関する状況 >

【サマリー】

実数表

総アクセス数(A)	2459
総ビジター数(V)	400
総参照時間(T)	20.461

係数表

1ビジターあたりのアクセス数(A/V[人])	6.147
1アクセスあたりの参照時間(T[分]/A)	0.499
1ビジターあたりの参照時間(T[分]/V[人])	3.069

曜日別実数表

曜日	アクセス数(A)	ビジター数[人](V)	参照時間[時間](T)
月曜	500	82	3.873
火曜	697	106	4.572
水曜	499	82	5.557
木曜	135	25	1.769
金曜	440	66	3.036
土曜	124	22	1.379
日曜	64	17	0.275

曜日別係数表

曜日	(A/V[人])	(T[分]/A)	(T[分]/V[人])
月曜	6.098	0.465	2.834
火曜	6.575	0.394	2.588
水曜	6.085	0.668	4.066
木曜	5.400	0.786	4.247
金曜	6.667	0.414	2.760
土曜	5.636	0.667	3.761
日曜	3.765	0.258	0.970

【時間帯別アクセス数】

時間帯	アクセス数	アクセス数/ビジター数	参照時間[分]/アクセス数
00:00~	146	9.125	0.474
01:00~	52	3.714	0.344
02:00~	63	4.846	0.749
03:00~	30	5.000	1.141
04:00~	22	2.200	0.129

⋮
⋮
⋮

【図12】

< 遷移ランキング >

【指定ページ(TOP)からの導線ランキング】
(上位3位)

ビジター数	コンテンツ
=====	
[指定ページ1]	/xxx/campaign1.html [キャンペーン1]

2852	/xxx/campaign1.html /index3.html /index4.html /index5.html /campaign/info.html

2000	/xxx/campaign1.html /index3.html /index4.html /index5.html /campaign/info2.html

1000	/xxx/campaign1.html /index2.html /index4.html /index5.html /info.html
=====	
[指定ページ2]	/xxx/campaign2.html

2823	/xxx/campaign2.html /test4.html /index4.html /index5.html /sample212.html

2000	/xxx/campaign2.html /index3.html /index4.html /index5.html /campaign/info2.html

1000	/xxx/campaign2.html /index2.html /info.html /index4.html /index5.html
=====	

【図14】

<u>顧客、サーバ情報テーブル</u>	<u>担当者テーブル</u>	<u>アクセス情報テーブル</u>
顧客ID	担当者ID	アクセス日時
レポート単位ID	氏名	参照元FQDN
レポート名	所属	参照元サブドメイン
顧客サーバID	e-mail	参照元フルドメイン
顧客サーバ名	WEB用ID	参照先URL
OS	WEB用PWD	リファラ
FTP		検索キーワード
FTP-ACCOUNT		同一アクセスフラグ
週次		
月次		
送信先アドレス		
サービス開始日		
サービス終了日		

<u>サービス提供者マスタ</u>	<u>ロボットマスタテーブル</u>	<u>顧客別サービス情報テーブル</u>
サービス提供者ID	レポート単位ID	レポート単位ID
サービス提供者名	収集範囲	分析種別
FROM	適用開始日	サイクル
Reply-to	適用終了日	出力順位
Errors-to	プロトコル	指定ページ最大数
担当者連絡先	ポートNo	遷移最大数
	コンテンツのパス	登録日
	データ1	
	データ2	
	データ3	
	データ4	
	データ5	
	種別	
	登録日	